

# Van der Waerden's construction of a splitting field

Fred Richman  
Florida Atlantic University

10 February 2005

## Abstract

In his classic book, *Modern Algebra*, van der Waerden gave a procedure for factoring polynomials over a finite-dimensional, separable, simple extension field. I believe that there is a nonconstructive component to his proof, and I will indicate where it comes in and why. Although I'm sure that this component could be avoided while staying within the framework that he set down, it is simpler to get around the problem by working with a splitting *algebra*, which is easily constructed for any polynomial and base field. The existence of a splitting field then follows from van der Waerden's argument.

## 1 Introduction

Van der Waerden was quite interested in developing algebra from a constructive point of view. In the preface to the second edition of *Modern Algebra* [5], he says:

“I have tried to avoid as much as possible any questionable set-theoretic reasoning in algebra. Unfortunately, a completely finite presentation of algebra, avoiding all non-constructive existence proofs, is not possible without great sacrifices. Essential parts of the algebra would have to be eliminated, or the theorems would have to be formulated with so many restrictions that the text would become unpalatable and certainly useless for a beginner.

On the other hand, it was possible at least to compile the building stones for a constructive foundation of algebra insofar as they exist at this time. In the theory of fields I did this by presenting the field-theoretical operations in a finite number of steps in such fashion that the intuitionistic foundation of the theory, insofar as it is possible today, can be seen readily.”

For an example of such an unpalatable text that is useless for a beginner, see [3].

This note deals with van der Waerden’s procedure for factoring polynomials over a finite-dimensional, separable, simple extension field, the main result in Section 42, *The field-theoretical operations in a finite number of steps*. An immediate corollary to this result is a procedure for constructing splitting fields of separable polynomials. The principal tool is the field norm, which is defined in three different ways in Section 41. The first way presupposes the existence of a splitting field so is unsuitable for his development. Although I’m sure that one could work directly with one of the other two definitions, I don’t believe van der Waerden did this. In the last section, I will indicate how to get around the problem by defining the norm in terms of the splitting *algebra* of a polynomial, which is easily constructed for any polynomial and base field. The existence of a splitting *field* then follows from van der Waerden’s argument.

## 2 Factorial fields

A field  $\Delta$  is said to be **factorial** [3] if you can factor any nonconstant polynomial in  $\Delta[X]$  into irreducible polynomials. Isn’t every field factorial? Van der Waerden [6] constructed a field  $\Delta$  of algebraic numbers and showed, in effect, that if you could factor any nonconstant polynomial in  $\Delta[X]$  into irreducible factors, then you could determine whether any given Turing machine halted. So you certainly couldn’t program a Turing machine to factor polynomials in  $\Delta[X]$ . His field  $\Delta$  was infinite dimensional over the rational numbers  $\mathbf{Q}$ , but you can get a feel for the idea behind his construction by simply thinking of a field  $\Delta$  between  $\mathbf{Q}$  and  $\mathbf{Q}[i]$  such that you don’t know whether  $i \in \Delta$ . Without that information, you will clearly be unable to factor the polynomial  $X^2 + 1$  into irreducible factors over  $\Delta$ .

Kronecker proved that  $\mathbf{Q}$  is a factorial field. Donald Knuth has pointed

out that Friedrich Schubert developed this technique 100 years before Kronecker. For a more complete history of this theorem, see [2]

Van der Waerden [5, Section 42] notes that if you want to construct a splitting field for a polynomial  $\varphi$  over a field  $\Delta$ , you run into the problem of factoring polynomials in extensions of  $\Delta$ . He goes on to show how to construct a splitting field for any separable irreducible polynomial  $\varphi$  provided that the base field  $\Delta$  is factorial. The hypothesis that  $\varphi$  be irreducible is inessential if the base field is factorial. What about the hypothesis that  $\varphi$  be separable? For a Brouwerian counterexample showing that you cannot omit this hypothesis, see [3, Exercise 5, Section VII.1]

### 3 The norms

Let  $\Delta$  be a field and  $\Sigma$  a finite-dimensional extension field of  $\Delta$ . In Section 41, van der Waerden gives three definitions of the norm  $N : \Sigma \rightarrow \Delta$ , one using conjugates, one using minimal polynomials, and one using determinants. We will denote these norms by  $N_1$ ,  $N_2$ , and  $N_3$ . For  $N_1$ , he assumes that  $\Sigma$  is a simple extension  $\Delta(\theta)$  where  $\theta$  satisfies an irreducible polynomial  $\varphi$  of degree  $n$  over  $\Delta$ . He writes  $\varphi(X) = (X - \theta_1) \cdots (X - \theta_n)$  over a splitting field of  $\varphi$  and defines the norm of an element

$$\eta = a_0 + a_1\theta + \cdots + a_{n-1}\theta^{n-1}$$

to be the product  $N_1(\eta) = \eta_1 \cdots \eta_n$  of the conjugates

$$\eta_i = a_0 + a_1\theta_i + \cdots + a_{n-1}\theta_i^{n-1}.$$

He observes that  $N_1$  can be extended in a natural way to a map from  $\Delta(\theta)[X_1, \dots, X_k]$  to  $\Delta[X_1, \dots, X_k]$ . This approach presupposes the existence of a splitting field for  $\varphi$ .

The  $N_2$  we compute the minimal polynomial  $g(X) = X^m + b_{m-1}X^{m-1} + \cdots + b_0$  over  $\Delta$  of the element  $\eta$  of  $\Delta(\theta)$ . This is unproblematic because  $\Delta(\theta)$  is finite dimensional over  $\Delta$ . Note that  $n = mr$  where  $r$  is the dimension of  $\Delta(\theta)$  over  $\Delta$ . Van der Waerden defines the  $N_2(\eta)$  to be  $(-1)^n b_0^r$ . He shows that  $N_2(\eta) = N_1(\eta)$ , given the existence of a splitting field for  $\varphi$ , by considering  $G(X) = N_1(X - \eta)$ . This is a polynomial of degree  $n$  over  $\Delta$  whose constant term is  $(-1)^n N_1(\eta)$ . Using the fact that  $g$  is the minimal polynomial of  $\eta$ , and of all the  $\eta_i$ , van der Waerden shows that  $G(X) = g(X)^r$ , so  $(-1)^n N_1(\eta) = b_0^r$ , which shows that  $N_1(\eta) = N_2(\eta)$ .

For  $N_3$ , van der Waerden drops the assumption that  $\Sigma$  is a simple extension. Multiplication by  $\eta$  gives a linear transformation of the finite-dimensional vector space  $\Sigma$  into itself, and  $N_3(\eta)$  is defined to be the determinant of that transformation. Van der Waerden gives the standard argument that the determinant is independent of the basis of  $\Sigma$  over  $\Delta$ . Now if  $g(X) = X^m + b_{m-1}X^{m-1} + \dots + b_0$  is the minimal polynomial of  $\eta$  over  $\Delta$ , as before, then by picking a basis for  $\Sigma$  that consists of multiples of the basis  $1, \eta, \dots, \eta^{m-1}$  of  $\Delta(\eta)$ , we see that  $N_3(\eta) = ((-1)^m b_0)^r = N_2(\eta)$ . So  $N_2$  and  $N_3$  are equal, and neither presupposes a splitting field.

Van der Waerden also notes that we may extend  $N_3$  in a natural way to a map  $\Sigma[X_1, \dots, X_k] \rightarrow \Delta[X_1, \dots, X_k]$ . I believe what he says here amounts to considering  $\Sigma[X_1, \dots, X_k]$  as a finite-rank free module over  $\Delta[X_1, \dots, X_k]$  and taking the determinant of the linear transformation induced by multiplication by an element of  $\Sigma[X_1, \dots, X_k]$ . It is clear that  $N_3$  is multiplicative.

Van der Waerden says at the end of Section 42 that, “The above proof is set up in such a way as not to presuppose the existence of a splitting field.” So he cannot use  $N_1$  but must use  $N_2$  or  $N_3$ . The danger here is the temptation to *assume* that properties of  $N_1$  are also properties of  $N_3$ . I think that’s what happened, but there is no way to be sure because it is always possible to argue that he had some other justification in mind.

Why would that be so bad? After all,  $N_1$  takes the same values as  $N_3$ , so they must have the same properties. Van der Waerden’s construction makes no reference to a splitting field. But that’s not what he says. He says that the *proof* does not presuppose the existence of a splitting field, and the proof that  $N_3$  has all the properties of  $N_1$  surely does. I’m quite confident that, unlike many computational algebraists, van der Waerden’s intent here was not only to provide an algorithm, but also to provide a *constructive* proof that it worked. That is the only way that “the intuitionistic foundation of the theory . . . can be seen readily.”

## 4 The factoring algorithm

The setting in Section 42 for the main construction is a factorial field  $\Delta$ , a monic separable irreducible polynomial  $\varphi$  of degree  $n$  over  $\Delta$ , and a root  $\theta$  of  $\varphi$  in an extension field of  $\Delta$ . Van der Waerden wants to show that the field  $\Delta(\theta)$  is also factorial, so he gives a procedure for factoring a monic polynomial  $f(Z)$  in  $\Delta(\theta)[Z]$ . He lets  $U$  be another indeterminate, and

considers

$$Nf(Z - U\theta) = F(Z, U) \in \Delta[Z, U]$$

“according to Section 41.” Presumably we should interpret this  $N$  as being  $N_3$ . He says that  $F(Z, U)$  is monic, which I guess means that it is monic as a polynomial in  $Z$  over the ring  $F[U]$ . That’s obvious for  $N = N_1$ . I suppose you could show it for  $N = N_3$ , but van der Waerden offers no justification. Personally, I believe that he is thinking of  $N_1$ .

Now he factors  $F(Z, U)$  into irreducible polynomials in  $\Delta[Z, U]$ . He can do this because  $\Delta$  is factorial, and if you can factor polynomials in one variable over  $\Delta$ , you can do it in many variables by a trick of Kronecker’s (which van der Waerden gives in this section). Taking the greatest common divisor, in the gcd-domain  $\Delta(\theta)[Z, U]$ , of  $f(Z - U\theta)$  with each of the irreducible factors of  $F(Z, U)$ , we find that one of those irreducible factors must divide  $f(Z - U\theta)$  for otherwise  $f(Z - U\theta)$  would be relatively prime to  $F(Z, U)$  when in fact it divides  $F(Z, U)$ . (The fact that  $\eta$  divides  $N_3\eta$  is proved in Section 41 as one of the “two more theorems, which we shall need in the following section”.) So either  $f(Z - U\theta)$  has a proper factorization in  $\Delta(\theta)[Z, U]$ , or it divides some  $\Delta[Z, U]$ -irreducible factor of  $F(Z, U)$ .

The map taking  $p(Z, U)$  to  $p(Z - U\theta, U)$  is an automorphism of  $\Delta(\theta)[Z, U]$ . Thus if  $f(Z - U\theta)$  has a proper factorization in  $\Delta(\theta)[Z, U]$ , then so does  $f(Z)$ , and those factors will have to lie in  $\Delta(\theta)[Z]$ . So either  $f(Z)$  factors in  $\Delta(\theta)[Z]$ , or  $f(Z - U\theta)$  divides some  $\Delta[Z, U]$ -irreducible factor of  $F(Z, U)$ . To finish the proof, van der Waerden shows that the latter alternative implies that  $f(Z)$  is irreducible in  $\Delta(\theta)[Z]$ .

So suppose that  $f(Z - U\theta)$  divides some  $\Delta[Z, U]$ -irreducible factor  $F_1(Z, U)$  of  $F(Z, U)$ , and that

$$f(Z) = f_1(Z) f_2(Z)$$

where  $f_1$  and  $f_2$  are monic polynomials in  $\Delta(\theta)[Z]$ . Replace  $Z$  by  $Z - U\theta$  and take norms to get

$$F(Z, U) = Nf_1(Z - U\theta) Nf_2(Z - U\theta).$$

So one of  $Nf_1(Z - U\theta)$  and  $Nf_2(Z - U\theta)$ , say  $Nf_1(Z - U\theta)$ , is divisible by  $F_1(Z, U)$  and therefore by  $f(Z - U\theta)$ . Write

$$Nf_1(Z - U\theta) = f(Z - U\theta) g(Z, U)$$

with  $g(Z, U)$  in  $\Delta(\theta)[Z, U]$ . Let  $m_1$  and  $m$  denote the degrees of the monic polynomials  $f_1$  and  $f$ . At this point van der Waerden considers the leading

terms in  $Z$  and  $U$  on both sides (the sum of the monomials of the largest total degree in the two variables  $Z$  and  $U$ ). For the left hand side he says that this term is  $N(Z - U\theta)^{m_1}$ . (There is a misprint here, with  $m$  and  $m_1$  interchanged.) Again, that's clear for  $N = N_1$ . How clear is it for  $N = N_3$ ? The claim seems to be that for any polynomial, the leading term of the norm is the norm of the leading term. Is that clear for  $N_3$ ? I have no doubt that you can prove it, but I don't see how to right off.

Continuing with the proof, on the right hand side the leading term is of the form  $(Z - U\theta)^m g_1(Z, U)$ . There is no problem with that. As  $m > m_1$ , the norm of  $Z - U\theta$  must be divisible by  $(Z - U\theta)^2$ . Substituting  $U = 1$  we see that  $N(Z - \theta)$  is divisible by  $(Z - \theta)^2$ . But  $N(Z - \theta) = \varphi(Z)$  which has no multiple factors. That's the desired contradiction. So  $f(Z)$  is irreducible in  $\Delta(\theta)[Z]$ .

Finally, to get the splitting field of  $\varphi$ , we simply keep adjoining roots of  $\varphi$  to  $\Delta$ , using the fact that at each step the extension field that we have so far is factorial, so we can find an irreducible factor which allows us to construct a simple extension field containing a root of the that factor.

## 5 Splitting algebras

Instead of using a splitting field of  $\varphi$  to define  $N_1$  on  $\Delta(\theta)$  we can just as well use any commutative  $\Delta$ -algebra over which  $\varphi$  splits. The virtue of this is that there is a natural, unproblematic construction of such a  $\Delta$ -algebra, so we can dispense with  $N_2$  and  $N_3$  and simply use  $N_1$  in the proof that  $\Delta(\theta)$  is factorial.

Let  $\varphi(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_0$  be a polynomial over the field  $\Delta$ . Let  $\sigma_i$  be the  $i$ -th elementary symmetric function in the indeterminants  $\theta_1, \dots, \theta_n$ , and let  $I$  be the ideal of  $\Delta[\theta_1, \dots, \theta_n]$  generated by the  $n$  elements  $\sigma_i - a_i$ . The **splitting algebra** for  $\varphi$  over  $\Delta$  is defined to be  $R = \Delta[\theta_1, \dots, \theta_n]/I$ , (see [1, 4]). It has the universal property that if  $C$  is any commutative algebra over  $\Delta$  in which the polynomial  $\varphi$  can be written as  $\varphi(X) = (X - c_1) \cdots (X - c_n)$ , then there is a unique  $\Delta$ -algebra map from  $R$  to  $C$  taking  $\theta_1, \dots, \theta_n$  to  $c_1, \dots, c_n$ .

We need to know that  $R$  is not the trivial ring, so that  $\Delta$  can be considered as a subring of  $R$ . (By the universal property, that's the same as saying that there is some nontrivial  $\Delta$ -algebra over which  $\varphi$  splits.) So we need to show that the ideal  $I$  is proper. One way to see this is to note

that  $\Delta[\theta_1, \dots, \theta_n]$  is an integral extension of  $\Delta[\sigma_0, \dots, \sigma_{n-1}]$  and the elements  $\sigma_0 - a_0, \dots, \sigma_{n-1} - a_{n-1}$  are algebraically independent generators of  $\Delta[\sigma_0, \dots, \sigma_{n-1}]$ , hence generate a proper (maximal) ideal of  $\Delta[\sigma_0, \dots, \sigma_{n-1}]$ .

The splitting algebra  $R$  need not be a field. Indeed, if  $R$  is a field, then the Galois group of  $\varphi$  is the full symmetric group on  $n$  letters because any permutation of  $\theta_1, \dots, \theta_n$  induces an automorphism of  $R$ .

Map  $\Delta[T]$  to  $R$  by taking  $f(T)$  to  $f(\theta_1)f(\theta_2)\cdots f(\theta_n)$  modulo  $I$ . The claim is that this map respects the equivalence on  $\Delta[T]$  given by the principal ideal generated by  $\varphi(T)$ , and that the image of the map lies in  $\Delta$ . Thus the map induces the desired map  $N_1 : \Delta(\theta) \rightarrow \Delta$ . For the first part of the claim, suppose  $f = g + h\varphi$ . The image of  $f$  in  $R$  is

$$(g(\theta_1) + h(\theta_1)\varphi(\theta_1))(g(\theta_2) + h(\theta_2)\varphi(\theta_2))\cdots(g(\theta_n) + h(\theta_n)\varphi(\theta_n)).$$

But  $\varphi(\theta_i) \in I$  because  $\theta_i$  satisfies the polynomial  $T^n + \sigma_{n-1}T^{n-1} + \cdots + \sigma_0 = (T - \theta_1)\cdots(T - \theta_n)$  over  $\Delta[\theta_1, \dots, \theta_n]$ . So the image of  $f$  in  $R$  is the same as the image of  $g$  in  $R$ . The second part of the claim is clear because the elementary symmetric polynomials in  $\Delta[\theta_1, \dots, \theta_n]$  are congruent modulo  $I$  to elements of  $\Delta$ .

Note that  $\varphi$  need not be irreducible here, although if irreducibility is not required, then we must replace  $\Delta(\theta)$  by  $\Delta[\theta]$  which need not be a field.

## References

- [1] BOURBAKI, NICOLAS, *Éléments de mathématique, Algèbre, Chapitres 4 à 7*, Masson, Paris, 1981.
- [2] MIGNOTTE, MAURICE, AND DORU ȘTEFĂNESCU, La première méthode générale de factorisation des polynômes. Autour d'un mémoire de F. T. Schubert, *Rev. Histoire Math.* **7** (2001), no. 1, 67–89.
- [3] MINES, RAY, FRED RICHMAN, AND WIM RUITENBURG, *A course in constructive algebra*, Springer-Verlag 1988.
- [4] POHST, MICHAEL, AND HANS ZASSENHAUS, *Algorithmic algebraic number theory*, Encyclopedia of mathematics and its applications, Cambridge University Press, 1989.
- [5] VAN DER WAERDEN, B.L., *Modern Algebra*, Frederick Ungar Publishing Co., New York 1953.

- [6] \_\_\_\_\_, Eine Bemerkung über die Unzerlegbarkeit von Polynomen, *Math. Annalen* **102** (1930) 738–739.